

# Interactive segmentation of abdominal aortic aneurysms in CTA images

Marleen de Bruijne, Bram van Ginneken, Max A. Viergever, and Wiro J. Niessen

## ABSTRACT

A model-based approach to interactive segmentation of thrombus in abdominal aortic aneurysms from CTA data is presented. After manual segmentation of the first slice, the method automatically detects the contour in subsequent slices, using the result from the previous slice as a reference. If an obtained contour is not sufficiently accurate, the user can intervene and provide an additional manual reference contour.

The method is inspired by the active shape model (ASM) segmentation scheme proposed by Cootes and Taylor,<sup>1</sup> in which a statistical shape model, derived from corresponding landmark points in manually labeled training images, is fitted to the image in an iterative manner. In our method, a shape model of the contours in two adjacent image slices is progressively fitted to the entire volume. The contour obtained in one slice constrains the possible shapes in the next slice. The optimal fit is determined on the basis of multi-resolution gray level models constructed from gray value patches sampled around each landmark. We propose to use the similarity of adjacent image slices for this gray level model, and compare these to single-slice features that are more generally used with ASM. The performance of various image features is evaluated in leave-one-out experiments on 23 data sets.

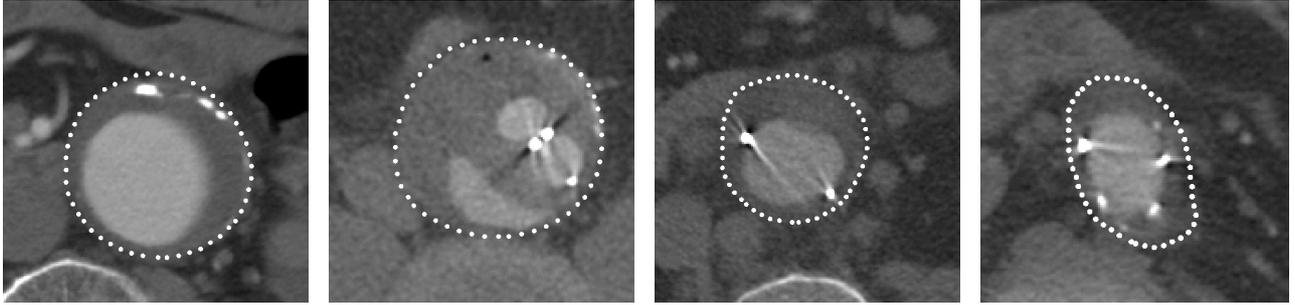
Features that use the similarity of adjacent image slices outperform measures based on single-slice features in all cases. The average number of slices in our datasets is 51, while on average 8 manual initializations are required, which decreases operator segmentation time by a factor of 6.

## 1. INTRODUCTION

An abdominal aortic aneurysm (AAA) is an enlargement of the infrarenal abdominal aorta, resulting from weakened arterial walls. Once present, AAAs continue to enlarge and become increasingly susceptible to rupture, which usually results in death. Because of the risks associated with surgical AAA treatment, patients with a small AAA are often placed under frequent surveillance until the aneurysm becomes symptomatic or its diameter exceeds 5.5 cm. Worldwide, approximately 100,000 surgical interventions for AAA repair are performed each year, of which at present 30% are endovascular. After endovascular aneurysm repair, in which a synthetic graft is placed inside the aorta, the process of aneurysm shrinkage, ongoing aneurysmal disease, and damage or fatigue of graft material may result in leakage, graft migration, and kinking or buckling of the graft, which can subsequently cause rupture or occlusion. Careful and frequent patient follow-up including regular CT scans will likely be needed for the life of the patient after endovascular aneurysm treatment.<sup>2</sup>

It has been demonstrated that change of aneurysm volume is a good indicator for the risk of aneurysm rupture.<sup>3-5</sup> Currently, the gold standard for volume assessment is Computed Tomography Angiography (CTA) with subsequent manual aneurysm delineation,<sup>6</sup> generally referred to as ‘thrombus segmentation’, since after successful graft placement the aneurysm sac surrounding the graft is completely filled with thrombus. The manual segmentation is a time-consuming process — it takes an experienced operator around 30 minutes — and suffers from inter- and intra-operator variations. Wever et al.<sup>7</sup> reported an inter-operator reproducibility coefficient of 8.3% for thrombus volume measurements.

To reduce analysis time and to increase reproducibility, automated segmentation would be of great value. However, thrombus segmentation in CTA images is a difficult task. The boundary can be obscured by surrounding tissue of similar gray value, many neighboring structures induce strong edges in close proximity to the aneurysm wall, and the variable aneurysm radius in combination with partial volume effects makes the boundary hard to distinguish even in the absence of neighboring structures. In addition, the size and appearance of aneurysms and neighboring structures vary considerably between patients as well as in one patient over time. Figure 1 shows some examples of CTA slices, illustrating the diversity of aneurysm and background appearance.



**Figure 1.** Four CTA slices of different patients, with manually delineated aneurysms. The left image is taken before surgery, the others are post-operative. The interior of the contour contains thrombus, contrast enhanced blood in the lumen and possibly in endoleaks, calcification, and — post-operatively — metal graft markers and attachment systems. Neighboring structures are the spinal column, at the bottom of the images, blood vessels and intestines. All CTA images in this thesis have been contrast stretched for better visibility of the thrombus.

Most publications on computerized AAA segmentation have concentrated on segmentation of the contrast-filled lumen, either in pre-operative<sup>8–18</sup> or in post-operative CTA scans.<sup>19</sup> The more difficult problem of thrombus segmentation has less frequently been addressed. Wilson et al.<sup>20</sup> reported successful thrombus segmentation using an active surface with the central lumen axis as a seed, but they did not present an evaluation study. Their work mainly aimed at providing realistic visualizations suitable for treatment planning. Subramanyan et al.<sup>21</sup> segmented the aortic lumen using a three-dimensional level-set approach. They propose to re-initialize the level-set in the image with the lumen subtracted to obtain the thrombus region, but present results on lumen diameter measurements only. Subasic et al.<sup>22</sup> also used a three-dimensional level-set for lumen segmentation, but found that the method could not cope with the large regions lacking boundary evidence in thrombus segmentation. They solved this problem using a combination of thresholding, edge detection, morphological operations, and shape restrictions in the form of a distance map, favoring convex cross-sections. A level-set was evolved in the resulting binary image to ensure a smooth object. The average segmentation error per slice was 19.8%, which makes the method suitable for visualization purposes but not for accurate volume measurements.

We present and evaluate an interactive method for AAA thrombus segmentation. After manual segmentation of the first slice, the method automatically detects the contour in subsequent slices, using the result from previous slices as a reference. If an obtained contour is not correct, the user can intervene and provide an additional manual reference contour. The method relies on the fitting of a shape model to points with high correlation with the reference contour.

The method is inspired by the Active Shape Model (ASM) segmentation scheme as put forward by Cootes and Taylor,<sup>1, 23</sup> which combines statistical knowledge of object shape and shape variation with local appearance models near object contours. ASMs have been successfully applied to various segmentation tasks in medical imaging.<sup>24–31</sup> Applications in three-dimensional images have either used a full three-dimensional model<sup>24, 32, 33</sup> or a two-dimensional model applied to image slices.<sup>25, 29</sup> We have chosen a slice-by-slice approach, which has the advantage that it is easily integrated with the current manual segmentation scheme. Moreover, the high anisotropy of the CTA data makes in-slice grey value models more appropriate, and a three-dimensional model would require resampling to an equal number of slices for all aneurysms, while in our datasets an aneurysm is between 30 and 65 slices long.

Mitchell et al.<sup>29</sup> proposed the use of a two-dimensional hybrid active shape model/active appearance model, fitted to several cardiac MR image slices independently. Cootes et al.<sup>25</sup> reported successful slice-by-slice ASM segmentation of the ventricles in 3D MR datasets, using the fitted shape in one slice as the initialization in the next slice. We extend this scheme by modeling two successive image slices together. The shape obtained in the previous slice is not only used as an initialization but constrains the shape in the next slice, resulting in a segmentation where the transition between slices is in accordance with the shapes in the training set.

Conventional ASMs use a linear model generated from gray value patches in training images to fit the shape

model to the image. AAAs have such a diverse range of possible surrounding structures, which vary in location, shape, intensity and texture, that a linear model of gray level structure about the contours may not be able to find the true contour in a new image. An important indicator for the image structure in one image slice is the structure in the adjacent slices. We propose a modified ASM scheme in which optimal landmark positions are defined by maximum gray value correlation with adjacent slices rather than by correlation with the training data. In this study, several similarity measures are evaluated and compared to features that were shown to perform well in combination with ASMs.<sup>1,31</sup>

The shape model is briefly discussed in Section 2.1. Section 2.2 presents various features for aneurysm boundary localization, and Section 2.3 explains how the shape model is fitted to the image using these features. Section 3 describes a series of leave-one-out experiments on 23 datasets. Several robustness tests are presented in Sections 3.4 and 3.5. The performance of the interactive procedure is measured by the required amount of user interaction in a simulated interactive segmentation scheme, as presented in Section 3.6. A discussion and conclusions are given in Sections 4 and 5.

## 2. METHODS

This section describes the segmentation scheme. Shape variations in the training set are described using a Point Distribution Model (PDM).<sup>34</sup> The shape model is used to generate new shapes, similar to those found in the training set, which are fitted to the data using a model of local gray value structure.

### 2.1. Point Distribution Models

In PDMs, a statistical model of object shape and shape variation is derived from a set of  $s$  training examples. Each training example is described by  $n$  manually or automatically identified corresponding landmark points. Variations in the coordinates of these landmark points describe the variation in shape and pose across the training set.

A shape is described by its shape vector  $\mathbf{x}$  containing all landmark coordinates, in 2D:

$$\mathbf{x} = (x_1, y_1, x_2, y_2, \dots, x_n, y_n)^T. \quad (1)$$

The shapes must be in a common coordinate frame. To maximize the specificity of the model, the shapes can be aligned by rotation, translation, and scaling, thus reducing non-linearities in the shape distribution. However, any relation between the object's shape and its pose or scale is then lost. The most popular approach for alignment of shapes is Procrustes Analysis,<sup>35</sup> which performs a least-squares minimization of the distances to the mean shape.

Principal Component Analysis (PCA) is applied to the shape vectors. To this end, the mean shape  $\bar{\mathbf{x}}$ , the covariance matrix  $\mathbf{S}$ , and the eigensystem of  $\mathbf{S}$  are computed:

$$\bar{\mathbf{x}} = \frac{1}{s} \sum_{i=1}^s \mathbf{x}_i \quad (2)$$

$$\mathbf{S} = \frac{1}{s-1} \sum_{i=1}^s (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T. \quad (3)$$

The eigenvectors  $\phi_i$  of  $\mathbf{S}$  provide the modes of shape variation present in the data. Any shape  $\mathbf{x}$  in the dataset can be represented as

$$\mathbf{x} = \bar{\mathbf{x}} + \mathbf{P}\mathbf{b}, \quad (4)$$

where  $\mathbf{P}$  is the concatenation of eigenvectors of  $\mathbf{S}$ , i.e.  $\mathbf{P} = (\phi_1 | \phi_2 | \dots | \phi_{2n})$ . The eigenvectors corresponding to the largest eigenvalues  $\lambda_i$  account for the largest shape variation; a small number of modes usually explains most of the variation. Each shape  $\mathbf{x}$  in the set can thus be approximated by

$$\mathbf{x} \approx \bar{\mathbf{x}} + \Phi\mathbf{b}, \quad (5)$$

where  $\Phi$  consists of the eigenvectors corresponding to the  $t$  largest eigenvalues,  $\Phi = (\phi_1|\phi_2|\dots|\phi_t)$ . The  $t$ -dimensional model parameter vector  $\mathbf{b}$  weighs the contribution of each of the eigenvectors,

$$\mathbf{b} = \Phi^T(\mathbf{x} - \bar{\mathbf{x}}). \quad (6)$$

The number of modes in the model,  $t$ , can be chosen in several ways. For instance, a threshold can be set on the reconstruction error that is allowed if the model is fitted to new shape instances. The minimum number of modes required can then be determined in cross-validation experiments. Another option is to estimate the variance of the noise on the landmark positions,  $\sigma_n^2$ , and retain all eigenvectors of which the corresponding eigenvalue exceeds the noise level,  $\lambda_i > \sigma_n^2$ .<sup>23</sup> Note, that this leads to an overestimation of the number of modes, owing to chance fluctuations in the limited sample. Paulsen et al.<sup>33</sup> therefore propose to use Horn’s parallel analysis<sup>36</sup> to determine the number of modes. In this analysis, the data is compared to samples of random vectors with the same variance and the same sample size. The eigenvalues of the data are plotted in decreasing order, and the same is done for the eigenvalues of the random data, averaged over all samples. The number of significant modes is given by the intersection point of the two curves. The dimensionality of the data can also be estimated using Bayesian model selection.<sup>37</sup>

The simplest and most widely used approach is to choose  $t$  such that the model explains at least a proportion  $f_v$  of the total variance observed in the training set. The total variance is given by  $\sum \lambda_i$ . We then select the  $t$  largest eigenvalues so that

$$\sum_{i=1}^t \lambda_i \geq f_v \sum_{i=1}^{2n} \lambda_i. \quad (7)$$

This is the approach we shall take in this chapter.

## 2.2. Gray value model

A rectangular gray value patch of  $k_n \times k_t$  pixels is sampled symmetrically around each landmark, with the length  $k_n$  normal to the contour and the width  $k_t$  tangential to the contour. The patches are sampled at multiple resolutions, to enable coarse-to-fine fitting of the model. The finest resolution uses the original image and a sample spacing of one voxel, the next resolution is the image convolved with a Gaussian kernel of width one voxel and sampled with a spacing of two voxels, and subsequent resolutions are obtained by doubling both the kernel width and the sample spacing.

For a successful segmentation procedure, the ability of an image feature to distinguish between correct and false boundary positions is crucial. We evaluate the performance of the following features.

### Edges

An object edge likely corresponds to image positions with a strong intensity gradient. We applied the gradient magnitude in the direction normal to the contour,  $GM$ .

### Linear model of training data

The boundary appearance can be learned from a training set. In the original ASM formulation, Cootes and Taylor<sup>1,23</sup> use the Mahalanobis distance from the sampled gray value patch  $\mathbf{g}_s$  to the mean of training patches  $\bar{\mathbf{g}}$  as a boundary measure:

$$MAH(\mathbf{g}_s) = (\mathbf{g}_s - \bar{\mathbf{g}})^T \mathbf{S}_g^{-1} (\mathbf{g}_s - \bar{\mathbf{g}}). \quad (8)$$

Minimizing  $MAH(\mathbf{g}_s)$  is equivalent to maximizing the probability that  $\mathbf{g}_s$  originates from the same distribution as the training patches. The effect of global intensity changes can be reduced by sampling the first derivative and normalizing the patch. We have applied  $MAH$  to intensity ( $I$ ) and first order derivative ( $I'$ ) patches, and the normalized versions thereof. The patches are normalized such that the sum of the absolute intensity values equals 1.

### Linear model of training data combined with slice similarity

The linear model of training patches that is used in ASM can be combined with information from adjacent slices. The Mahalanobis distance from the sample patch  $\mathbf{g}_s$  to the reference patch  $\mathbf{g}_r$  expresses the difference between the two patches, taking the covariance across the training set into account:

$$MAHR(\mathbf{g}_s) = (\mathbf{g}_s - \mathbf{g}_r)^T \mathbf{S}_g^{-1} (\mathbf{g}_s - \mathbf{g}_r). \quad (9)$$

The reference patch is sampled at the corresponding landmark in the adjacent slice. In the experiments, *MAHR* is applied to (normalized) intensity and first order derivative patches.

### Slice similarity

Several features based only on similarity between adjacent slices are considered:

- Sum of squares of intensity differences between sample and reference

$$SD = \sum_{x=0}^{k_n \times k_t} (I_s(x) - I_r(x))^2 \quad (10)$$

where  $I_s$  is the image intensity in the sample patch and  $I_r$  the intensity in the reference patch.

- Sum of absolute intensity differences between sample and reference

$$AD = \sum_{x=0}^{k_n \times k_t} |I_s(x) - I_r(x)| \quad (11)$$

- Normalised cross correlation between sample and reference

$$NCC = \frac{\sum_{x=0}^{k_n \times k_t} I_s(x) \cdot I_r(x)}{\sqrt{\sum_{x=0}^{k_n \times k_t} I_s(x)^2 \cdot \sum_{x=0}^{k_n \times k_t} I_r(x)^2}} \quad (12)$$

- Correlation coefficient between sample and reference

$$CC = \frac{\sum_{x=0}^{k_n \times k_t} (I_s(x) - \bar{I}_s) \cdot (I_r(x) - \bar{I}_r)}{\sqrt{\sum_{x=0}^{k_n \times k_t} (I_s(x) - \bar{I}_s)^2 \cdot \sum_{x=0}^{k_n \times k_t} (I_r(x) - \bar{I}_r)^2}} \quad (13)$$

All slice similarity features are tested on raw intensity patches.

*GM*, *NCC*, and *CC* are maximized, all other measures are minimized.

### 2.3. Fitting procedure

Starting with one manually drawn contour, the segmentation proceeds through the dataset, fitting the two-slice model to all pairs of adjacent slices. Each time the result from the previous slice is used as a reference for determining the contour in the new slice. From coarse to fine resolution, the model is fitted to a pair of slices in an iterative procedure. The process initializes with the shape from the reference slice, in both slices. For all landmarks in the new slice,  $n_s$  possible new positions are evaluated along the line perpendicular to the contour, on both sides. The distance between the positions to evaluate depends on the level of resolution and is equal to the sample spacing of the corresponding gray level model: at the finest resolution the distance is 1 voxel, at each subsequent resolution it is doubled. The optimal new position is determined by one of the features that were introduced in the previous section. The shape model parameters  $\mathbf{b}$  that minimize the squared distance between

the landmarks and the optimal positions are computed using Equation 6. Hard limits are applied to constrain  $\mathbf{b}$  to plausible values:

$$|b_i| \leq f_c \sqrt{\lambda_i}. \quad (14)$$

The landmarks of the new slice are moved to their new positions according to the model parameters, while the landmarks in the reference slice are kept fixed. This process is repeated a fixed number of  $N$  times, whereupon it is repeated at the next level of resolution.

### 3. EXPERIMENTS

A series of leave-one-out experiments is performed on 23 CTA images.

#### 3.1. Data

All experiments are carried out on 23 routinely acquired CTA images from 23 different patients, including 2 pre-operative and 9 post-operative scans. The remaining 12 scans are taken at follow-up ranging from 1 to 12 months. The scan resolution is  $0.488 \times 0.488 \times 2.0$  mm. Each image consists of circa 125 slices of  $512 \times 512$  voxels, of which 34 to 63 slices contain aneurysmal tissue. The total number of slices to be segmented is 1175.

#### 3.2. Landmarking and alignment

A PDM is built of all pairs of adjacent slices. The original CT slices, which are perpendicular to the body axis and therefore always give approximately perpendicular cross-sectional views of the aorta, are used. Contours are drawn manually by an expert. In the absence of well-defined anatomical landmark points, landmarks are equidistantly placed along the object contours. The starting point of a contour is the posterior point with the same  $y$ -coordinate as the center of mass. All pairs of contours are aligned such that the centers of the reference contours coincide. Note, that no size normalization is performed, since a correlation between shape and size is expected.

Figure 2 shows the first eight modes of shape variation of the two-slice model trained with all 23 datasets. The largest variation in the shapes in the training set, the first mode, is a size difference. The second and third mode show that a strong elongation in  $x$  direction is correlated with a slightly larger upper contour, while an elongation in the  $y$  direction corresponds to a smaller upper contour. These are the types of shapes that are found in the bottom and the top of the aneurysm. The next two modes describe a less pronounced deviation from the mean shape, where in the fourth mode both slices deviate in the same way and in the fifth mode there is a small difference between both slices. Only from the fifth mode, a larger difference between the two adjacent slices is observed, with first a shape difference, then translations in the  $x$  and  $y$  directions, followed by a difference in size.

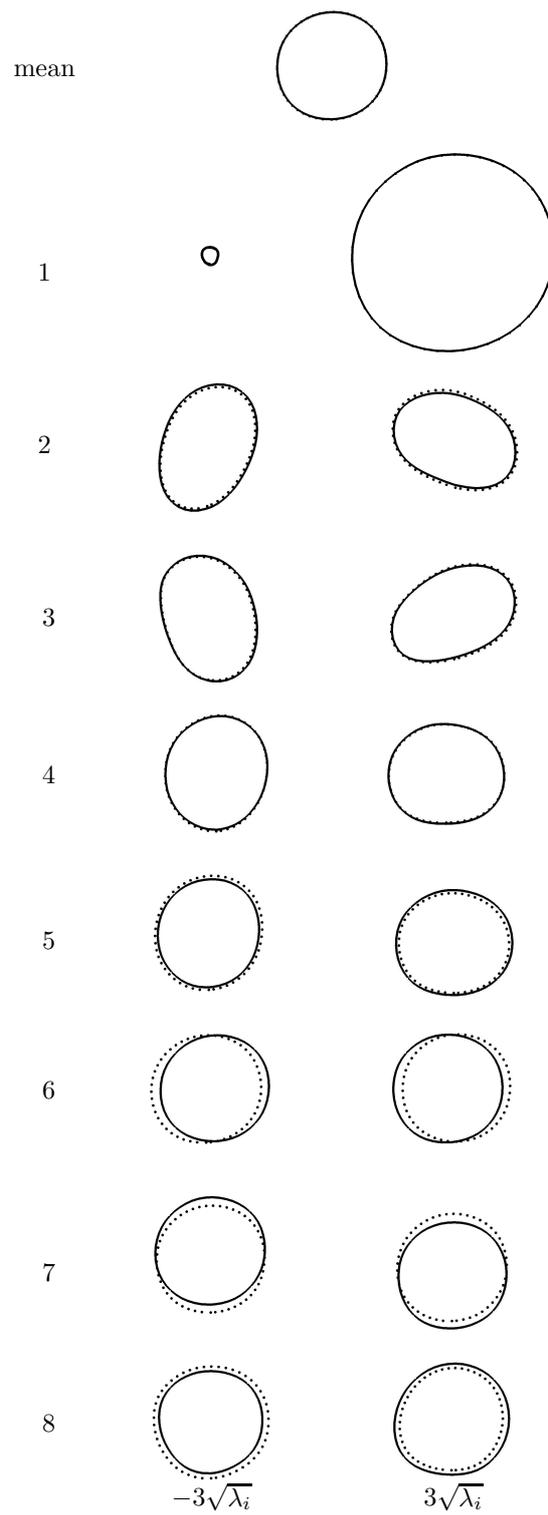
#### 3.3. Parameter settings

Table 1 lists the parameter values that are used in all experiments, unless mentioned otherwise.

#### 3.4. Shape model evaluation

The amount of variance explained by the model is plotted as a function of the number of modes in Figure 3a. Over 99% of the total amount of variation present in the training set is described by the first 11 modes.

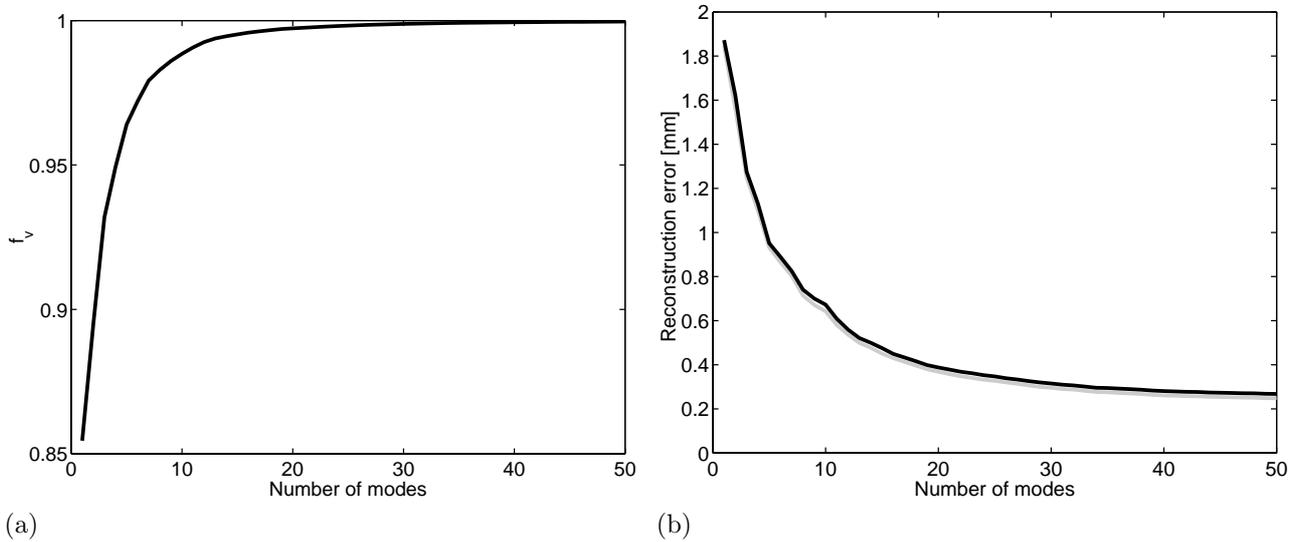
The validity of the shape model and the applied constraints are tested by fitting the model directly to the contours that were manually drawn, which gives an upper bound for the accuracy that can be obtained when the model is fitted to new image data. Figure 3b shows the root mean squared error in millimeters as a function of the number of modes retained. The averaged leave-one-out reconstruction error, where a model built of 22 training shapes is fitted to the remaining shape, almost coincides with the reconstruction error for a model built of all 23 shapes, indicating that sufficient training data are available to represent all shapes.



**Figure 2.** The effect of varying each of the first eight shape parameters individually in a two-slice model built from all 23 datasets. The continuous contour denotes the lower slice, the dotted contour the upper slice.

**Table 1.** Parameters of the segmentation scheme.

Shape model (Section 2.1)		
$n$	50	Number of landmark points per slice
$f_v$	0.99	Part of shape variance to be explained by the model, controlling the number of modes $t$ (Equation 7)
$f_c$	2.5	Bound on eigenvalues $\lambda_i$ (Equation 14)
Gray value model (Section 2.2)		
$k_n$	7	Patch length
$k_t$	1	Patch width
Fitting algorithm (Section 2.3)		
$n_s$	2	Number of new landmark positions to evaluate on either side of the current position
$L$	3	Number of resolution levels
$N$	5	Number of iterations per resolution level



**Figure 3.** (a) The amount of variance explained by the model as a function of the number of modes contained in the model. (b) Root mean squared error as a function of the number of modes, for a model built of all datasets (gray line) and the leave-one-out case (black line).

### 3.5. Feature evaluation

First, the features' capability to find the true contour is evaluated independent of the rest of the segmentation process. Starting from the landmarks on the manually drawn contour, the optimal landmark positions according to the feature are determined, for varying  $n_s$ . The distances from these points to the landmark positions provided by the manual tracings are measured. The resulting root mean squared distances as a function of the size of the search region are shown in Figure 4. Features using patches of length  $k_n = 7$  and width  $k_t = 1$  and  $k_t = 5$ , with a sample spacing of 1 or 4 voxels, are considered. Initial experiments revealed that smaller errors are obtained with the Mahalanobis distance based measures when the patches are normalized. For the sake of clarity of the graph, we here only show the results on normalized intensity and derivative patches, but the final results on non-normalized patches are given in Section 3.6 for comparison.

At a small scale and in a small search region, all features perform similar, but at larger scale or in a larger region the errors obtained using slice similarity measures are much smaller than those obtained using the gradient magnitude or the linear statistical model. Both Mahalanobis distance based features perform much better on derivative than on intensity patches. The Mahalanobis distance to the average intensity patch is even worse than random landmark selection. Among the slice similarity features, *AD* and *SD* yield better results than *NCC* and *CC*. The robustness of the slice similarity features increases with increasing scale, while the robustness of the Mahalanobis distance and gradient measures decreases. The same holds for enlarging the patch width.

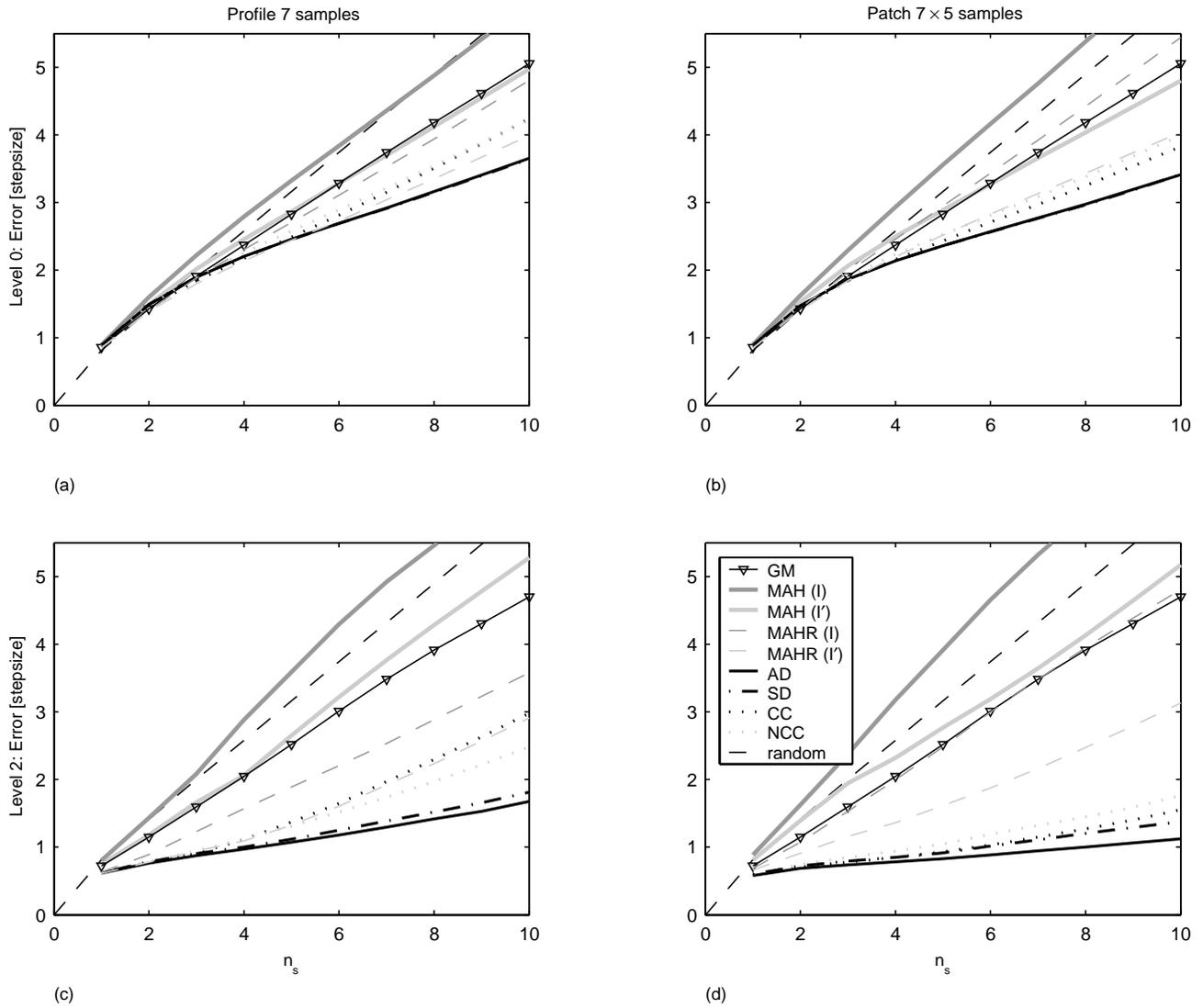
The segmentation process is less sensitive to random outliers than to several neighboring landmarks that are consistently placed at a wrong position, which can be the case when a neighboring structure produces misleading boundary evidence. The result of fitting the shape model to the optimal points of the previous test is given in Figure 5. The minimum error has increased, since it cannot be smaller than the reconstruction error from Figure 3.b, and all maximum errors have decreased. Again, both Mahalanobis distance based features perform better on derivative than on intensity patches. The Mahalanobis distance to the reference patch gives errors similar to those obtained using the other slice similarity measures. *GM* is one of the best features if the search regions is within 3 voxels of the true contour, but errors quickly increase with larger search areas. At the lowest resolution, *NCC* and *CC* perform better than *AD* and *SD*, but at higher resolutions *AD* gives the smallest errors. Enlarging the patch width does not increase the robustness for the slice similarity measures as was the case before fitting the shape model, and robustness decreases for all Mahalanobis distance based measures. In the subsequent experiments we use  $k_t = 1$ .

### 3.6. Simulated interactive segmentation

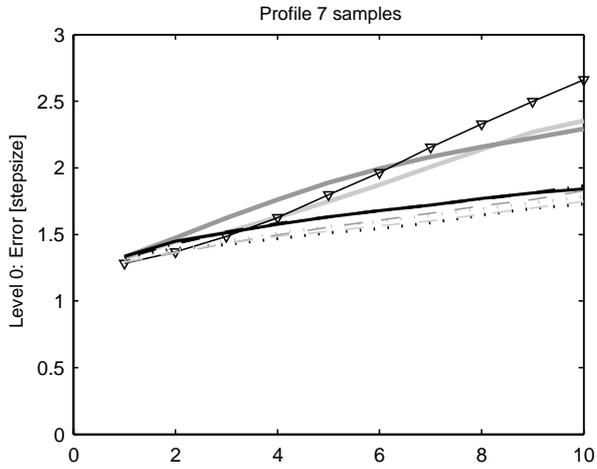
To evaluate all features evenly in an interactive segmentation setting, we simulated observer interaction using manually segmented datasets. Starting with the bottom contour of the manual segmentation as a reference, successive slices are segmented. After each segmented contour, the similarity to the manually drawn shape is computed. If the similarity is sufficient, the segmentation proceeds, otherwise, a new contour from the available manual segmentation is added, so as to simulate operator intervention. The performance is measured by the number of manual initializations needed to segment an entire volume.

The similarity of the two contours is described in terms of the distance from each landmark to the contour drawn manually. Since the main clinical objective of AAA segmentation is volume measurement, the mean signed distance, which is proportional to the error in the measured area in a slice, should be small (1 mm in our experiments). To ensure similarly shaped and positioned contours the mean and maximum absolute distance are constrained as well, to 2 mm and 4 mm respectively. Four randomly chosen examples of segmentations fulfilling these requirements are shown in Figure 6.

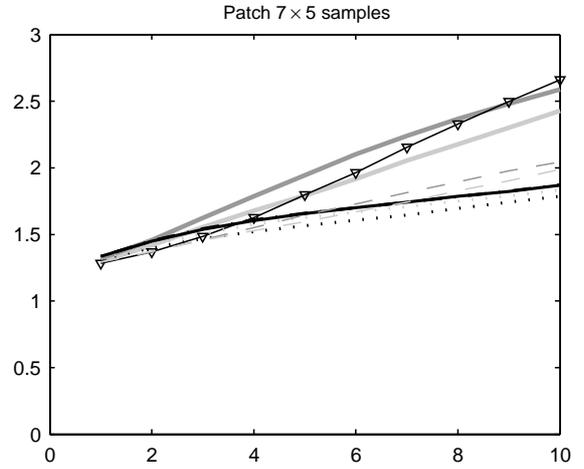
Figure 7 shows, for each feature and each dataset, the number of slices that was correctly segmented using one initialization. Table 2 lists the number of correctly segmented slices using one initialization, averaged over all datasets. In all cases, the smallest number of interactions was needed when one of the slice similarity measures was used. The average number of automatically obtained contours was largest when using *AD*, but the difference with *SD* and *CC* was not significant ( $p = 0.75$  and  $p = 0.06$  in a paired t-test). *AS*, *SD* and *CC* performed significantly better than all other tested features ( $p < 0.00001$ ). If *MAHR* was used, thus combining the information from the reference slice with the statistical model, more user interactions were needed, and using only the model (*MAH*) again increased the number of interactions ( $p = 0.003$ ). Overall, Mahalanobis distance



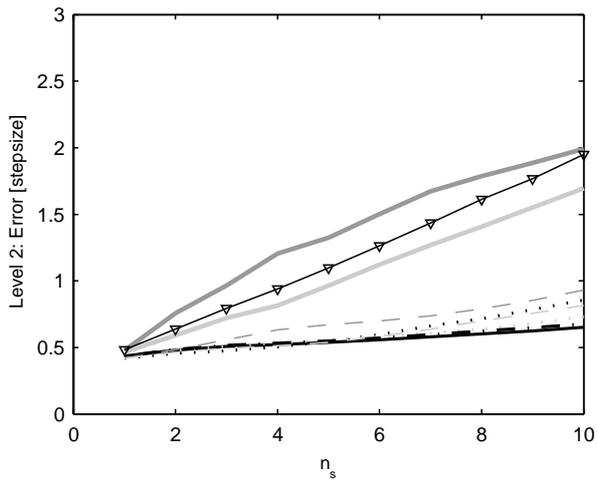
**Figure 4.** Root mean squared error of optimal landmark position detection as a function of the number of points evaluated on either side of the contour, for (a) a patch of  $7 \times 1$  voxels, sampled at scale 0; (b) a patch of  $7 \times 5$  voxels, sampled at scale 0; (c) a patch of  $7 \times 1$  voxels, sampled at scale 2; (d) a patch of  $7 \times 5$  voxels, sampled at scale 2. The units are step sizes in the position evaluation procedure, i.e. 1 voxel (0.488mm) for the profiles sampled at scale 0 and 4 voxels at scale 2. The dashed black line starting at the origin corresponds to the expectation for random landmark selection.



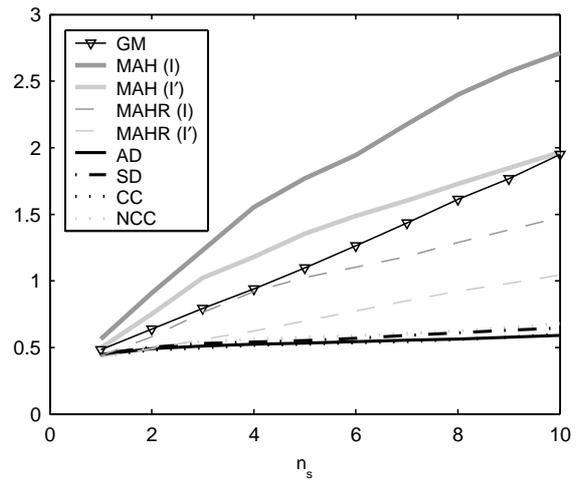
(a)



(b)

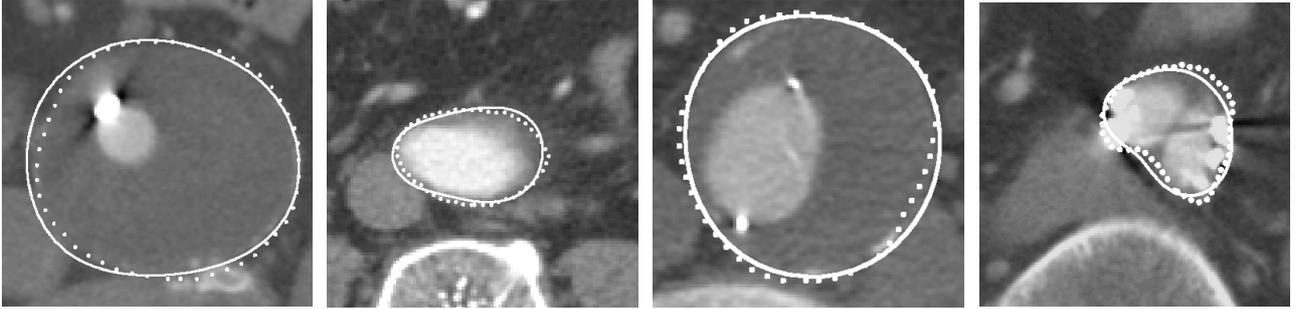


(c)



(d)

**Figure 5.** Root mean squared error of landmark positions as a function of the number of points evaluated on either side of the contour, after fitting the shape model to the optimal positions, for (a) a patch of  $7 \times 1$  voxels, sampled at scale 0; (b) a patch of  $7 \times 5$  voxels, sampled at scale 0; (c) a patch of  $7 \times 1$  voxels, sampled at scale 2; (d) a patch of  $7 \times 5$  voxels, sampled at scale 2. The units are step sizes in the position evaluation procedure, i.e. 1 voxel (0.488mm) for the profiles sampled at scale 0 and 4 voxels at scale 2.



**Figure 6.** Examples of accepted segmentations. The continuous line depicts the contour that was obtained automatically, the dotted line is the manual segmentation.

**Table 2.** Average number of correct contours obtained using one initialization.

	normalized	
<i>GM</i>	1.82	
<i>MAH(I)</i>	1.32	1.16
<i>MAH(I')</i>	1.11	2.25
<i>MAHR(I)</i>	2.86	2.96
<i>MAHR(I')</i>	2.61	3.23
<i>AD</i>	4.76	
<i>SD</i>	4.72	
<i>CC</i>	4.24	
<i>NCC</i>	3.59	
<i>AD, Supervised</i>	6.39	

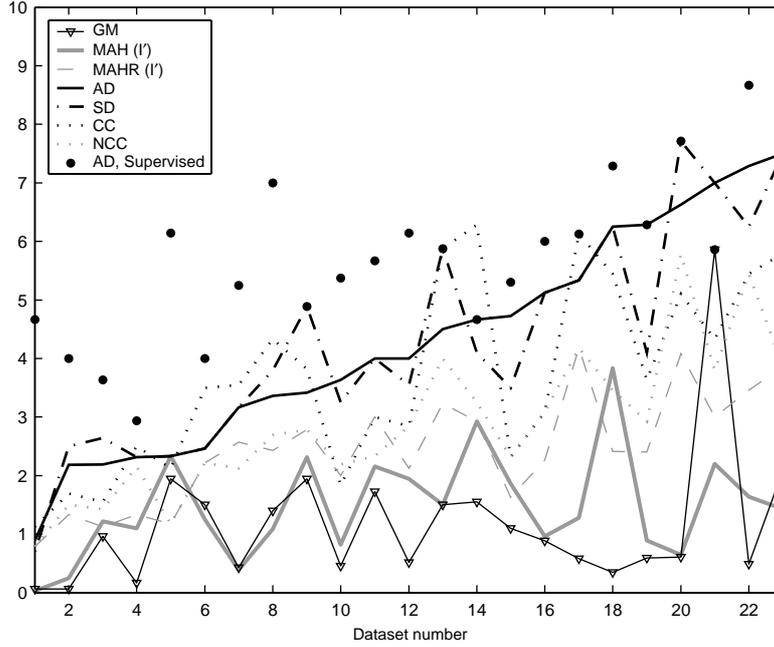
based measures perform best on normalized first order derivative patches. A segmentation process based on *MAH(I')* on normalized patches needs less re-initializations than one based on *GM* ( $p = 0.011$ ).

### 3.7. Expert supervised interactive segmentation

In the simulated experiments the manual tracings are used as the undisputed gold standard. However, inter-observer errors in aneurysm volume measurements are as high as 8.3%.<sup>7</sup> In case of clear image evidence the automatic segmentation should indeed be similar to the manual contour, but when no boundary is visible the criteria of Section 3.6 may be relaxed. To get an idea of the true number of interactions that will be needed for segmentation, segmentation with feature *AD* was repeated, but this time a human observer decided whether re-initialization was needed. This resulted in a smaller number of interactions, especially in the top and bottom part of the aneurysm where the boundary is hard to distinguish owing to partial volume effects. The number of interactions used is given in Figure 7 and Table 2. The resulting segmentations had an average volume overlap of 95.8% (minimum 93.0%) with the manual segmentation and an average relative volume difference of 1.5% (maximum 3.9%)

## 4. DISCUSSION

The presented method uses a shape model from two adjacent slices for segmenting a three dimensional structure, while a full three-dimensional approach — allowing gray value modeling perpendicular to the object’s surface — could be more appropriate in some applications. In the case of CTA images, which are in general highly anisotropic (in the images used in this study the voxels are over 4 times larger in the  $z$ -direction), we do not expect a significant improvement in boundary localization if the profiles would be sampled in three dimensions.



**Figure 7.** Average number of correct contours obtained with one initialization, shown for all datasets.

An advantage of the two-slice approach is that it can easily be integrated in the current clinical practice of manual slice-by-slice segmentation. With the current ASM optimization procedure, the use of slice similarity features, which turned out to be more accurate in aneurysm boundary localization, is computationally very expensive for a model consisting of a large number of slices. Moreover, the use of a three dimensional model would require resampling to an equal number of landmarks for all datasets, which is not desirable for objects that can vary largely in length.

It was shown that the original ASM scheme that uses local intensity models does not perform well in AAA segmentation. One of the shortcomings of the gray value model that is used in ASMs is that only the appearance of the correct boundary is learned from the training set, wherefore it cannot distinguish the true contour from other contours with similar gray values. Another problem is the underlying assumption of a normal intensity distribution. In the presence of distinct background structures of varying shape and brightness this assumption does not hold. The use of a non-linear gray value model may be more suitable. In this work we effectively solved this problem by using gray value patches from the reference slice.

A consequence of using slice similarity measures is that errors propagate through the dataset and therefore performance deteriorates with distance from the reference contour. We previously showed that a refinement step, in which the model is allowed to adjust to nearby edges, can increase contour tracking robustness<sup>38</sup> in the middle part of the aneurysm. However, the model can then also fit to wrong edges. Figure 5 shows that the errors quickly become large when fitting on edge strength. In segmentation of the entire aneurysm, including the more difficult top and bottom part, we have not observed a significant improvement when such a refinement was applied.

The proposed method could obtain accurate segmentations while only in one of six slices manual intervention is needed. Currently, the user must redraw the entire contour, but more sophisticated methods of user interaction can easily be incorporated. Since errors often occur at a small part of the contour while the larger part is correct, dragging one landmark point to the correct boundary and iteratively fitting the model to this point may be sufficient to obtain a correct contour.<sup>39</sup>

We have evaluated the application of this method to AAA segmentation, but the idea of propagating a two-

slice shape model based on slice similarity measures can be used in other areas as well. Especially time-sequences and anisotropic medical images form a promising application. The dependency on image slicing and orientation restricts the applicability to segmentation tasks where the shape change between slices is not too large and the object is always imaged in approximately the same direction. However, many medical images are acquired with fixed scan protocols which satisfy these conditions. The proposed method could also be used for segmentation in reformatted slices orthogonal to a predefined object axis, which is a common approach in vessel segmentation.

## 5. CONCLUSIONS

A new approach to the semi-automatic segmentation of thrombus in abdominal aortic aneurysms, based on shape model fitting in sequential slices, is reported. In a simulated interactive segmentation process, the benefit of using slice correlation in shape model fitting was demonstrated. Using the proposed method and the optimal slice correlation feature, the time required for expert segmentation may be reduced by a factor of 6.

## REFERENCES

1. T. Cootes, C. Taylor, D. Cooper, and J. Graham, "Active shape models – their training and application," *Computer Vision and Image Understanding*, vol. 61, no. 1, pp. 38–59, 1995.
2. M. Fillinger, "Postoperative imaging after endovascular AAA repair," *Seminars in Vascular Surgery*, vol. 12, no. 4, pp. 327–338, 1999.
3. J. Wever, J. Blankensteijn, W. Mali, and B. Eikelboom, "Maximum aneurysm diameter follow-up is inadequate after endovascular abdominal aortic aneurysm repair," *European Journal of Vascular and Endovascular Surgery*, vol. 20, no. 2, pp. 177–182, 2000.
4. B. Czermak, G. Fraedrich, M. Schocke, I. Steingruber, P. Waldenberger, R. Perkmann, M. Rieger, and W. Jaschke, "Serial CT volume measurements after endovascular aortic aneurysm repair," *Journal of Endovascular Therapy*, vol. 8, no. 4, pp. 380–389, 2001.
5. J. Pollock, S. Travis, S. Whitaker, I. Davidson, R. Gregson, B. Hopkinson, P. Wenham, and S. MacSweeney, "Endovascular AAA repair: Classification of aneurysm sac volumetric change using spiral computed tomographic angiography," *Journal of Endovascular Therapy*, vol. 9, no. 2, 2002.
6. R. Balm, R. Kaatee, J. Blankensteijn, W. Mali, and B. Eikelboom, "CT-Angiography of abdominal aortic aneurysms after transfemoral endovascular aneurysm management," *European Journal of Vascular and Endovascular Surgery*, vol. 12, no. 2, pp. 182–188, 1996.
7. J. Wever, J. Blankensteijn, J. van Rijn, I. Broeders, B. Eikelboom, and W. Mali, "Inter- and intra-observer variability of CTA measurements obtained after endovascular repair of abdominal aortic aneurysms," *American Journal of Roentgenology*, vol. 175, no. 5, pp. 1297–1282, 2000.
8. M. Fiebich, M. Tomiak, R. Engelmann, J. McGilland, and K. Hoffman, "Computer assisted diagnosis in CT angiography of abdominal aortic aneurysms," in *Medical Imaging: Image Processing* (K. Hanson, ed.), vol. 3034 of *Proceedings of SPIE*, pp. 86–94, SPIE Press, 1997.
9. A. Bulpitt and E. Berry, "Spiral CT of abdominal aneurysms: comparison of segmentation with an automatic 3D deformable model and interactive segmentation," in *Medical Imaging: Image Processing* (K. Hanson, ed.), vol. 3338 of *Proceedings of SPIE*, pp. 938–946, SPIE Press, 1998.
10. A. Bulpitt, E. Berry, R. Boyle, D. Scott, and D. Kessel, "A deformable model, incorporating expected structure information, for automatic 3D segmentation of complex anatomical structures," in *Computer Assisted Radiology and Surgery* (H. Lemke, M. Vannier, K. Inamura, A. Farman, and K. Doi, eds.), vol. 1214 of *Excerpta Medica International Congress Series*, pp. 572–577, Elsevier Publishers, 2000.
11. G. Rubin, D. Paik, P. Johnston, and S. Napel, "Measurement of the aorta and its branches with helical CT," *Radiology*, vol. 206, pp. 823–829, Mar 1998.
12. O. Wink, W. Niessen, and M. Viergever, "Fast delineation and visualization of vessels in 3-D angiographic images," *IEEE Transactions on Medical Imaging*, vol. 19, no. 4, pp. 337–346, 2000.
13. H. Tek, D. Comaniciu, and J. Williams, "Vessel detection by mean shift-based ray propagation," in *IEEE Workshop on Mathematical Methods in Biomedical Image Analysis* (L. Staib and A. Rangarajan, eds.), IEEE Computer Society Press, 2001.

14. R. Pohle and K. Toennies, "A new approach for model-based adaptive region growing in medical image analysis," in *Computer Analysis of Images and Patterns* (W. Skarbek, ed.), vol. 2124 of *Lecture Notes in Computer Science*, pp. 238–246, Springer, 2001.
15. J. Weese, M. Kaus, C. Lorenz, S. Lobregt, R. Truyen, and V. Pekar, "Shape constrained deformable models for 3D medical image segmentation," in *Information Processing in Medical Imaging* (M. Insana and R. Leahy, eds.), vol. 2082 of *Lecture Notes in Computer Science*, pp. 380–387, Springer, 2001.
16. D. Magee, A. Bulpitt, and E. Berry, "Combining 3D deformable models and level set methods for the segmentation of abdominal aortic aneurysms," in *Proceedings of the British Machine Vision Conference* (T. Cootes and C. Taylor, eds.), pp. 119–126, The British Machine Vision Association, 2001.
17. K. Subramanyan, D. Smitha, J. Varmab, and S. Chandraa, "Automatic vessel extraction and abdominal aortic stent planning in multi-slice CT," in *Medical Imaging: Image Processing* (M. Sonka and M. Fitzpatrick, eds.), vol. 4684 of *Proceedings of SPIE*, pp. 114–122, SPIE Press, 2002.
18. A. Giachetti, M. Tuveri, and G. Zanetti, "Reconstruction and web distribution of measurable arterial models," *Medical Image Analysis*, vol. 7, no. 1, pp. 79–93, 2002.
19. M. de Bruijne, W. Niessen, J. Maintz, and M. Viergever, "Semi-automatic aortic endograft localisation for post-operative evaluation of endovascular aneurysm treatment," in *Medical Imaging: Image Processing* (M. Sonka and K. Hanson, eds.), vol. 4322 of *Proceedings of SPIE*, pp. 395–406, SPIE Press, 2001.
20. L. Wilson, S. Brown, J. Young, R. Li, and L. Brandt, "Three-dimensional computer models of abdominal aortic aneurysms by knowledge-based segmentation," in *Computer Assisted Radiology and Surgery* (H. Lemke, M. Vannier, K. Inamura, and A. Farman, eds.), vol. 1191 of *Excerpta Medica International Congress Series*, pp. 213–217, Elsevier Publishers, 1999.
21. K. Subramanyan, M. Steinmiller, D. Sifri, and D. Boll, "Automatic aortic vessel tree extraction and thrombus detection in multi-slice CT," in *Medical Imaging: Image Processing* (M. Sonka and M. Fitzpatrick, eds.), vol. 5032 of *Proceedings of SPIE*, SPIE Press, 2003. In Press.
22. M. Subasic, S. Loncaric, and E. Sorantin, "3D image analysis of abdominal aortic aneurysm," in *Medical Imaging: Image Processing* (M. Sonka and M. Fitzpatrick, eds.), vol. 4684 of *Proceedings of SPIE*, pp. 1681–1689, SPIE Press, 2002.
23. T. Cootes and C. Taylor, "Statistical models of appearance for computer vision," tech. rep., University of Manchester, 2002.
24. A. Hill, A. Thornham, and C. Taylor, "Model-based interpretation of 3D medical images," in *Proceedings of the British Machine Vision Conference* (J. Illingworth, ed.), pp. 339–348, The British Machine Vision Association, 1993.
25. T. Cootes, A. Hill, C. Taylor, and J. Haslam, "The use of active shape models for locating structures in medical images," *Image and Vision Computing*, vol. 12, no. 6, pp. 355–366, 1994.
26. S. Solloway, C. Hutchinson, J. Waterton, and C. Taylor, "The use of active shape models for making thickness measurements of articular cartilage from MR images," *Magnetic Resonance in Medicine*, vol. 37, no. 6, pp. 943–952, 1997.
27. N. Duta and M. Sonka, "Segmentation and interpretation of MR brain images: An improved active shape model," *IEEE Transactions on Medical Imaging*, vol. 17, no. 6, pp. 1049–1067, 1998.
28. G. Hamarneh and T. Gustavsson, "Combining snakes and active shape models for segmenting the human left ventricle in echocardiographic images," in *Proceedings IEEE Computers in Cardiology*, vol. 27, pp. 115–118, 2000.
29. S. Mitchell, B. Lelieveldt, R. van der Geest, H. Bosch, J. Reiber, and M. Sonka, "Multistage hybrid active appearance model matching: segmentation of left and right ventricles in cardiac MR images," *IEEE Transactions on Medical Imaging*, vol. 20, no. 5, pp. 415–423, 2001.
30. B. van Ginneken, A. Frangi, J. Staal, B. ter Haar Romeny, and M. Viergever, "Active shape model segmentation with optimal features," *IEEE Transactions on Medical Imaging*, vol. 21, no. 8, pp. 924–933, 2002.
31. G. Behiels, F. Maes, D. Vandermeulen, and P. Suetens, "Evaluation of image features and search strategies for segmentation of bone structures in radiographs using active shape models," *Medical Image Analysis*, vol. 6, no. 1, pp. 47–62, 2002.

32. M. Kohnen, A. Mahnken, J. Kesten, E. Koepfel, R. Günther, and B. Wein, "A three dimensional knowledge based surface model for segmentation of organic structures," in *Medical Imaging: Image Processing* (M. Sonka and M. Fitzpatrick, eds.), vol. 4684 of *Proceedings of SPIE*, pp. 485–494, SPIE Press, 2002.
33. R. Paulsen, R. Larsen, S. Laugesen, C. Nielsen, and B. Ersbøll, "Building and testing a statistical shape model of the human ear canal," in *Medical Imaging Computing & Computer-Assisted Intervention* (T. Dohi and R. Kikinis, eds.), vol. 2489 of *Lecture Notes in Computer Science*, Springer, 2002.
34. T. Cootes, D. Cooper, C. Taylor, and J. Graham, "Training models of shape from sets of examples," in *Proceedings of the British Machine Vision Conference* (D. Hogg and R. Boyle, eds.), pp. 9–19, Springer, 1992.
35. C. Goodall, "Procrustes methods in the statistical analysis of shape," *Journal of the Royal Statistical Society B*, vol. 53, no. 2, pp. 285–339, 1991.
36. J. Horn, "A rationale and test for the number of factors in factor analysis," *Psychometrika*, vol. 30, pp. 179–186, 1965.
37. T. Minka, "Automatic choice of dimensionality for PCA," Tech. Rep. 514, M.I.T. Media Laboratory Perceptual Computing Section, 2002.
38. M. de Bruijne, B. van Ginneken, W. Niessen, J. Maintz, and M. Viergever, "Active shape model based segmentation of abdominal aortic aneurysms in CTA images," in *Medical Imaging: Image Processing* (M. Sonka and M. Fitzpatrick, eds.), vol. 4684 of *Proceedings of SPIE*, pp. 463–474, SPIE Press, 2002.
39. B. van Ginneken, M. de Bruijne, M. Loog, and M. Viergever, "Interactive shape models," in *Medical Imaging: Image Processing* (M. Sonka and M. Fitzpatrick, eds.), vol. 5032 of *Proceedings of SPIE*, SPIE Press, 2003. In Press.